

# A Closed-Form, Analytical Approximation for Apparent Surface Charge and Electric Field of Molecules

Dan E. Folescu and Alexey V. Onufriev\*

Cite This: *ACS Omega* 2022, 7, 26123–26136

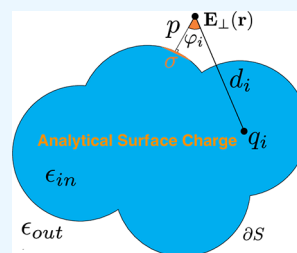
Read Online

ACCESS |

Metrics &amp; More

Article Recommendations

**ABSTRACT:** Closed-form, analytical approximations for electrostatic properties of molecules are of unique value as these can provide computational speed, versatility, and physical insight. Here, we have derived a simple, closed-form formula for the apparent surface charge (ASC) as well as for the electric field generated by a molecular charge distribution in aqueous solution. The approximation, with no fitted parameters, was tested against numerical solutions of the Poisson equation, where it has produced a significant speed-up. For neutral small molecules, the hydration free energies estimated from the closed-form ASC formula are within 0.8 kcal/mol RMSD from the numerical Poisson reference; the electric field at the surface is in quantitative agreement with the reference. Performance of the approximation was also tested on larger structures, including a protein, a DNA fragment, and a viral receptor–target complex. For all structures tested, a near-quantitative agreement with the numerical Poisson reference was achieved, except in regions of high negative curvature, where the new approximation is still qualitatively correct. A unique efficiency feature of the proposed “source-based” closed-form approximation is that the ASC and electric field can be estimated individually at any point or surface patch, without the need to obtain the full global solution. An open-source software implementation of the method is available: <https://people.cs.vt.edu/~onufriev/CODES/aasc.zip>.



## 1. INTRODUCTION

Accurate and efficient modeling of solvation effects at the atomistic level is a critical component of modern efforts to understand molecular structure and function.<sup>1–5</sup> Analysis and visualization of electrostatic properties of biomolecules, including the electric field and surface charge generated by the molecular charge distribution, have made an impact on qualitative reasoning about biomolecules.<sup>1,6</sup>

There are two broad approaches to the modeling of molecular electrostatics and solvation effects: explicit and implicit solvation methods.<sup>7</sup> Arguably the most widely used model of solvation is that for which individual solvent molecules are treated explicitly, on the same footing with the target molecule. However, accuracy of the explicit solvent representation comes at a high price computationally, limiting the practical utility of atomistic simulations in many areas. The implicit, continuum solvation approach—treating solvent as a continuum with the dielectric and nonpolar properties of water—can offer much greater effective simulation speeds compared to the explicit solvent models.<sup>8–18</sup> The Poisson equation<sup>9,10,19–23</sup> of classical electrostatics<sup>24</sup> provides an exact formalism—within the continuum, local, linear-response dielectric approximation of solvent in the absence of mobile ions—for computing the electrostatic potential  $V(\mathbf{r})$  produced by a molecular charge distribution  $\rho(\mathbf{r})$  characterizing the solute

$$\nabla \epsilon(\mathbf{r}) \nabla V(\mathbf{r}) = -4\pi \rho(\mathbf{r}) \quad (1)$$

where  $\epsilon(\mathbf{r})$  is the dielectric constant. Once  $V(\mathbf{r})$  is obtained, the electrostatic part of the solvation free energy is easily computed.<sup>24</sup>

The problem of finding  $V(\mathbf{r})$  is mathematically equivalent<sup>25,26</sup> to finding a continuous charge density,  $\sigma$ , on the dielectric boundary (DB), such that

$$V(\mathbf{r}) = \sum_i \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|} + \oint_{\partial S} \frac{\sigma(\mathbf{s})}{|\mathbf{r} - \mathbf{s}|} d^2s \quad (2)$$

where  $\rho(\mathbf{r})$  is the discrete charge distribution, formed by  $n$  point charges  $q_1, \dots, q_n$  and  $\sigma(\mathbf{s})$  is the apparent surface charge (ASC) associated with each surface patch  $\mathbf{s}$ . The second term in eq 2 represents the so-called reaction field potential.<sup>27–29</sup> Conceptually, once the ASC,  $\sigma(\mathbf{s})$ , is found, all of the solvation effects, at the level of the Poisson equation, can be computed. The reformulation of the Poisson problem via eq 2 has a number of technical advantages made apparent over the years, especially in quantum mechanical (QM) calculations.<sup>8,32</sup> A great variety of practical widely used methods, including multiple modern derivatives of PCM<sup>25</sup> and COSMO,<sup>30</sup> utilize

Received: March 12, 2022

Accepted: June 10, 2022

Published: July 19, 2022



this general idea—the apparent surface charge (ASC) formalism, see refs 31, 32 for comprehensive reviews.

A number of ASC-based methods yield numerically exact solutions to the Poisson equation, in the sense that the exact solution can, at least in principle, be approximated to arbitrary precision. Formally exact linear-scaling implementations of numerical ASC methods based on a conjugate gradient or domain decomposition exist.<sup>32</sup> Concerns related to the computational cost of numerically exact approaches<sup>31</sup> have led to the development of approximate ASC-based methods, such as the widely used COSMO,<sup>30</sup> GCOSMO,<sup>33</sup> and C-PCM.<sup>34</sup> These methods rely on approximations to eq 2.

Still, even these approximate ASC-based methods employ a significant numerical component, such as numerical matrix inversion, which may carry appreciable computational overhead, especially as the structure size grows.<sup>35,36</sup> Therefore, in applications where computational efficiency and algorithmic simplicity is paramount, numerical ASC-based methods may not be as competitive as approximations to the Poisson equation based purely on closed-form analytical expressions.<sup>27</sup>

Among the fully analytical approximations to the Poisson equation, the generalized Born (GB) model<sup>8,13,37–62</sup> is arguably the most widely used, especially in atomistic simulations.<sup>63–70</sup> However, despite its multiple documented success stories,<sup>71</sup> the GB model does not have the versatility of eq 2 and the associated benefits of an ASC-based formulation of biomolecular electrostatics. Here, we aim to fill the gap by deriving an analytical, closed-form approximation to the Poisson equation for the ASC and the (normal) electric field around an arbitrarily shaped molecule. Standard numerical solutions of the Poisson equation are used as the reference. We refrain from comparisons with well-established, optimized numerical implementations of ASC methods in this initial investigation.

The outline of the paper is as follows: Section 2 describes computational testing materials and methodology. Section 3 is focused on our analytical ASC approximation: we first derive an exact analytical ASC reference on a sphere (3.1) and present our approximate form of the ASC for arbitrary molecular geometries (3.2). We test the approximate ASC against the exact analytical ASC reference using a spherical test case (3.3), which simulates relevant electrostatic configurations that we will encounter in later sections. Application of our model to solvation energy calculations is presented in Section 3.4. Numerical performance and accuracy analysis along with an example applications are presented in Section 4, where we first examine the computational speed of our method (4.1) before testing its accuracy on small and large molecules (4.2.1 and 4.2.2). Finally, we showcase our analytical ASC approximation on a presently relevant biomolecular complex: that of the human ACE2 receptor and SARS-CoV-2 spike glycoprotein (4.3).

## 2. METHODS

**2.1. Structures.** For testing, a set of 173 neutral small molecules from version 0.52 of the FreeSolv database<sup>72,73</sup> was used. The original set of nearly 650 molecules was narrowed to include only those molecules containing hydrogen, oxygen, nitrogen, and carbon atoms. The small molecules under consideration are all rigid—having small conformational variability as seen in molecular dynamic (MD) simulations.<sup>74</sup> The choice of rigid molecules has allowed us to focus on the physics of solvation while mitigating the uncertainty related to

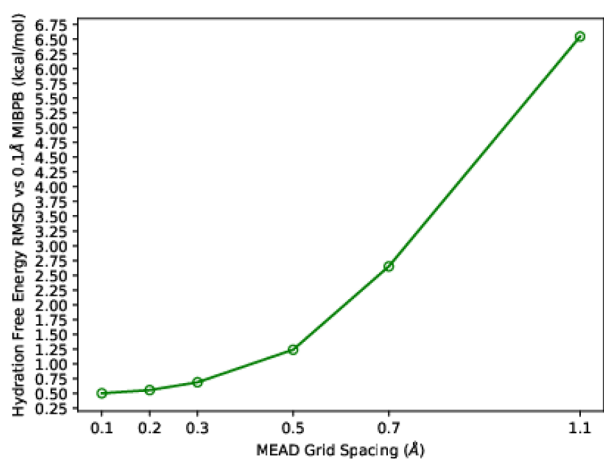
conformational sampling. *ambpdb*<sup>76</sup> was used to generate PQR format files from AMBER format coordinate and topology files.<sup>72</sup> Additionally, two larger biomolecules were used: a 25 bp poly-A B'-form dsDNA<sup>6</sup> and the hen-egg lysozyme (PDB: 2LZT).<sup>76</sup> Our method was also tested on a portion of the ACE2/SARS-CoV-2 complex (PDB: 6M0J)<sup>77</sup> receptor-binding domain (RBD). 6M0J RBD residues were determined through A/E chain contacts within 3.8 Å. In each chain, residues within 1.5 Å of the contacts were also included. The H++ server<sup>78</sup> was used to generate protonated PQR format files. PQR format files for small molecules and the SARS-CoV-2 complex along with RBD contact residue lists are provided in the accompanying code package.

**2.2. Dielectric Boundary Representations.** Just like the numerical Poisson solvers, ASC-based methods rely on molecular boundary representations.<sup>79–81</sup> These representations of the actual molecular shape are crucial to the accuracy and, to some extent, the efficiency of modern implicit solvation models. The question of which definition of the dielectric boundary (DB) is most appropriate is nontrivial; there is no universal default.<sup>82–84</sup> Here, our main goal is to assess the accuracy of the new ASC approximation against its primary accuracy metric, the numerical Poisson–Boltzmann (NPB). For this purpose, both methods can utilize whichever DB one considers most appropriate for the given application. Our secondary goal is to compare the new ASC approximation to a fast analytical GB model, which dictates the choice of the DB representation chosen here.

Within our ASC method, we approximate the solute–solvent interface—the DB—using the solvent excluded surface (SES),<sup>79,85</sup> with Bondi<sup>86</sup> atomic radii and a water probe of 1.4 Å. This DB is triangulated with the open-source package NanoShaper.<sup>80</sup> In each relevant test, the NanoShaper grid spacing was matched with the grid spacing used by the NPB reference; 0.1 Å was used for small-molecule accuracy comparisons, 0.25 Å for fair speed comparisons, and 0.5 Å for large-molecule electric field normal comparisons. The two existing reference methods employed in this work, the NPB and GB (see Sections 2.3 and 2.4), utilized matching dielectric boundaries based on the same set of Bondi radii with a 1.4 Å water probe radius (unless otherwise specified), for consistency.

**2.3. Numerical Poisson–Boltzmann Reference.** For NPB reference calculations, the macroscopic electrostatics with atomic detail (MEAD) package<sup>87</sup> was used. MEAD is a volumetric, finite-difference solver that can compute potential maps and hydration energies, utilizing an SES DB representation. The package was chosen primarily because it interfaces well with the visualization and analysis utility GEM,<sup>36</sup> used here to process the NPB-generated potential maps. Using GEM, visualization of the electric field around the solute can be achieved at any given distance from the DB. Convergence<sup>23</sup> of MEAD-generated hydration free energies was verified (Figure 1), supporting its usage as a reference when a fine enough grid resolution was taken. A modern, 2nd-order method MIBPB<sup>88</sup> was used as the accuracy reference for MEAD.

Based in part on the convergence analysis, potential-map calculations were performed at a grid spacing of 0.1 Å and hydration energy calculations at an inner and outer bounding-box grid spacing of 0.1 and 0.5 Å, respectively, for small-molecule accuracy comparisons. NPB reference solvation energy calculations utilize inner and outer dielectrics of 1



**Figure 1.** Convergence of hydration free energies computed by MEAD. The convergence is assessed against reference values computed by MIBPB,<sup>22</sup> as RMSD over our small molecule set. MIBPB hydration free energies are computed with a 1.4 Å water probe radius at a 0.1 Å grid spacing. MEAD hydration free energies are computed with identical parameters, except for the variable grid spacing. MIBPB small-molecule electrostatic hydration free energies ranged from  $-0.05$  to  $-14.37$  kcal/mol, with an average magnitude of 6.68 kcal/mol.

and 80, respectively. The water probe radius is 1.4 Å, unless otherwise stated. For fair speed comparisons, potential-map calculations were performed at a grid spacing of 0.25 Å and hydration energy calculations at a single bounding-box grid spacing of 0.25 Å, considered standard for finite-difference NPB calculations. For larger molecules, potential-map calculations were performed at a grid spacing of 0.5 Å and hydration energy calculations at an inner and outer bounding-box grid spacing of 0.5 Å and 1.0 Å, respectively. The total number of NPB grid points was determined by setting a volumetric bounding-box side length slightly larger (+1 Å) than the maximal intramolecular distance. The electric field normal at a point  $\mathbf{r}$  was numerically approximated by a two-point stencil:

$$\mathbf{E}_{\perp}(\mathbf{r}) = -\frac{\partial V}{\partial \hat{n}} \approx -\frac{V(\mathbf{r} + h\hat{n}) - V(\mathbf{r} - h\hat{n})}{2h} \quad (3)$$

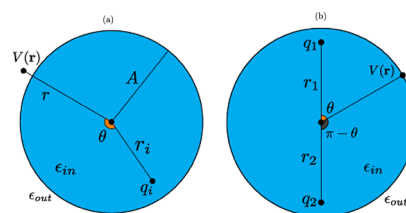
where  $\mathbf{r} + h\hat{n}$  and  $\mathbf{r} - h\hat{n}$  are two sampling point distance  $\pm h$  from the DB along the surface normal  $\hat{n}$ ; see ref 36 for additional details of the sampling protocol. Here,  $h$  was chosen to minimize the distance between sampled points, while still being large enough so that the sampled points are distinct. Notice that if  $h$  were too small, the sampling protocol would sample the exact same points of the cubic lattice used in the NPB reference calculations. The largest possible distance between two such grid points is the diagonal of the cubic grid, which determines the minimum  $h$ . To illustrate this numerical constraint,  $h$  must be larger than  $\sqrt{3}/(0.1)^2 \sim 0.173$  Å for members of the small molecule dataset, whose potential grids were computed with the NPB reference at a grid spacing of 0.1 Å.

Additionally, to avoid numerical artifacts of the NPB reference near the DB and mitigate possible effects of minor differences between the internal representations of the SES computed by our NPB reference and NanoShaper, the field computed a distance  $p > h$  from the DB (Figure 4): doing so ensures that we do not accidentally sample grid points inside

the molecule. The need to use a non-zero projection distance in the NPB calculations makes it necessary to consider the electric field normal values near the DB as the numerical reference for assessing the accuracy of the analytical ASC, rather than the apparent surface charge itself (which, up to a prefactor, is essentially the normal component of the field right at the DB, see eq 5).

**2.4. Generalized-Born Solvation Free Energy Reference.** The IGBS<sup>51</sup> GB model from AMBER<sup>75</sup> was utilized. This model was parameterized against reference NPB hydration free energies calculated based on the SES surface, Bondi radii, and a 1.4 Å water probe radius.

**2.5. Accuracy Metrics Used.** The ASC approximation was tested against the exact Poisson–Boltzmann (EPB, see Section 3.1) reference, with additional tests undertaken against the NPB and GB references. Our comparison with the EPB reference was used directly for the apparent surface charge, employing the two test charge configuration in Figure 2b. Per-



**Figure 2.** Geometry and charge settings for the Kirkwood multipolar expansion(a) and for the two-charge test case (b). These notations and geometries are used in eqs 4 and 6. In each case, a perfectly spherical, sharp DB is utilized, with  $\epsilon_{in}$  and  $\epsilon_{out}$  denoting inner and outer dielectric constants, respectively. The angle  $\theta$  (resp.  $\pi - \theta$ ) is subtended by the lines connecting the point of observation,  $\mathbf{r}$ , and the charge(s) to the spherical center. In panel (a), a single point charge,  $q_i$ , is located  $r_i$  away from the center of a spherical boundary with radius  $A$ . In panel (b), two point charges,  $q_1$  and  $q_2$ , are located on the vertical diameter of the sphere. The charges are of equal distance,  $r_1 = r_2$ , from the spherical center. The electrostatic potential  $V(\mathbf{r})$  is computed at the point of observation  $\mathbf{r}$ .

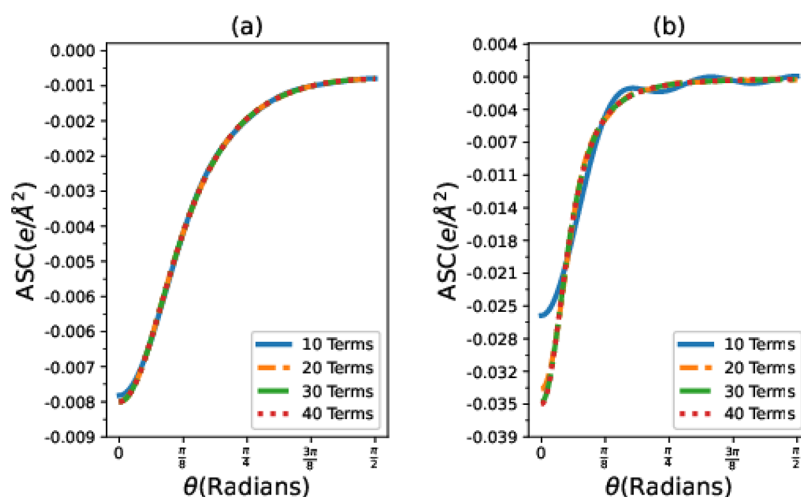
vertex electric field normal values were compared against the NPB reference, averaging over each vertex in a given biomolecule, and over each biomolecule in the comparison set. Electrostatic hydration free energies were compared against the NPB and GB references with inner and outer dielectrics  $\epsilon_{in} = 1$  and  $\epsilon_{out} = 80$ , unless otherwise specified. All results will be in  $e/\text{Å}^2$  for apparent surface charge, kcal/(mol  $\cdot e \cdot \text{Å}$ ) for electric field normals, and kcal/mol for electrostatic hydration free energies.

**2.6. Computer Specifications.** All computations and visualizations were completed on a commodity desktop computer with an Intel Core i7 (or equivalent) processor, using a maximum of 32 GB of memory.

### 3. THEORY AND RESULTS

**3.1. Preliminaries: Analytical Poisson–Boltzmann Reference.** In the context of implicit solvation, the simplest scenario is that of a solute with a sharp, spherical DB, Figure 2a.

For such a spherical boundary, as in Figure 2a, Kirkwood<sup>89</sup> gave the exact, analytical solution of eq 1 for the potential  $V_i$  at the DB due to a single charge  $q_i$  inside the boundary. Here, we use the solution valid on or exterior to the spherical DB,<sup>92</sup> without consideration for mobile ions. At  $r = A$



**Figure 3.** Apparent surface charge (ASC, per unit area) computed using the truncated, infinite series analytical solution of the Poisson problem on a sphere, eq 6. The convergence tests are conducted on the dual-positive test case, Figure 2b. The ASC is sampled at the spherical boundary  $A = 10$  Å away from the center an angle  $\theta$  from  $q_1$ . (a)  $r_1 = r_2 = 6$  Å; (b)  $r_1 = r_2 = 8$  Å; see Figure 2b. Partial sums of the infinite series solution, eq 6, with  $M = 10, 20, 30,$  and  $40$  terms are examined. The corresponding truncated sums are shown with blue, orange, green, and red lines, respectively.

$$V_i = -\frac{q_i}{A} \left( \frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \sum_{l=0}^{\infty} \left[ \frac{1}{1 + \left( \frac{l}{l+1} \right) \left( \frac{\epsilon_{\text{in}}}{\epsilon_{\text{out}}} \right)} \right] \left( \frac{r_i}{A} \right)^l P_l(\cos \theta) + \frac{q_i}{A} \left( \frac{1}{\epsilon_{\text{in}}} \right) \sum_{l=0}^{\infty} \left( \frac{r_i}{A} \right)^l P_l(\cos \theta) \quad (4)$$

The apparent surface charge  $\sigma$  is related to the normal component of the electric field  $\mathbf{E}_{\perp} = \left( \frac{\partial V}{\partial \vec{n}} \right)_{\text{out}}$  at (just outside) the boundary via<sup>24,25</sup>

$$\sigma = \frac{1}{4\pi} \left( 1 - \frac{\epsilon_{\text{out}}}{\epsilon_{\text{in}}} \right) \left( \frac{\partial V}{\partial \vec{n}} \right)_{\text{out}} \quad (5)$$

From this and eq 4, we obtain an exact, analytical expression for the apparent surface charge on the spherical DB:

$$\sigma_{\text{KW}} = \sum_i \frac{q_i}{4\pi} \left( \frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \left[ \sum_{l=0}^{\infty} \left[ \frac{1}{1 + \left( \frac{l}{l+1} \right) \left( \frac{\epsilon_{\text{in}}}{\epsilon_{\text{out}}} \right)} \right] \times (l+1) \left( \frac{r_i^l}{A^{l+2}} \right) P_l(\cos \theta) - \left( \frac{1}{\epsilon_{\text{in}}} \right) \sum_{l=0}^{\infty} (l+1) \left( \frac{r_i^l}{A^{l+2}} \right) P_l(\cos \theta) \right] \quad (6)$$

where the summation is over all of the enclosed charges. Equation 6 will provide a key check for our analytical ASC approximation.

**3.1.1. Convergence Analysis of  $\sigma_{\text{KW}}$ .** The presence of the indexing term,  $(l+1)$ , is notable in its effects on the convergence characteristics of eq 6, by increasing the number of terms necessary to obtain a converged, accurate reference. As the ratio  $r_i/A$  approaches 1—that is, as the charge approaches the DB—slow convergence of the approximate solution manifests itself.

In Figure 3b, we see that, even for  $\frac{r_i}{A} = 0.8$ , it is possible to achieve both qualitatively and quantitatively reasonable results. The analytical reference appears well-converged, Figure 3, for our purposes, at  $M = 30$  terms. Hence, eq 6 was numerically approximated by truncating to the first  $M = 30$  terms, calling the resulting expression the essentially exact Poisson–Boltzmann (EPB) reference.

**3.2. Main Result: Analytical Apparent Surface Charge.** To derive our ASC approximation, we begin with the previously derived closed-form approximation for the electrostatic potential around (outside) an arbitrary molecular shape,<sup>90</sup> see Figure 4:

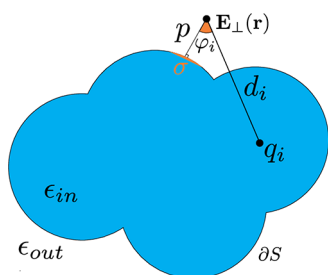
$$V_i \approx \left( \frac{q_i}{\epsilon_{\text{out}} \left( 1 + \alpha \left( \frac{\epsilon_{\text{in}}}{\epsilon_{\text{out}}} \right) \right)} \right) \left[ \frac{(1 + \alpha)}{d_i} - \frac{\alpha \left( 1 - \frac{\epsilon_{\text{in}}}{\epsilon_{\text{out}}} \right)}{r} \right] \quad (7)$$

One significant drawback of the approximation in eq 7 that will be overcome here is that it only provides a simple expression for the electrostatic potential in the outside (solvent) space. Thus, the computation of many quantities of interest remain out of reach, e.g., of the solvation free energy, which requires knowledge of the potential inside the solute.

To proceed, we utilize the polar orthonormal frame,  $e_r = \frac{\partial}{\partial r}$ ;  $e_{\theta} = \frac{1}{r} \frac{\partial}{\partial \theta}$ , to take its derivative, for use in eq 5. The derivative vanishes in the direction of  $e_{\theta}$ , yielding

$$\mathbf{E}_{\perp}(\mathbf{r}) = -\frac{\partial V_i}{\partial \vec{n}} = -\left[ \frac{\partial V_i}{\partial r} - \cos(\varphi_i) \frac{\partial V_i}{\partial d_i} \right] \quad (8)$$

Exploiting the geometry in Figure 4, we relate  $\cos(\varphi_i)$  as a dot product of the surface unit normal,  $\hat{n}$ , and the vector from  $\mathbf{E}_{\perp}(\mathbf{r})$  to  $q_i$ , which we denote  $\vec{d}_i$ :  $\cos(\varphi_i) = (\hat{n} \cdot \vec{d}_i)/d_i$ . Applying eq 7 to eq 8 and summing over the charge distribution (Figure 4), we arrive at:



**Figure 4.** Two dielectric problem for an arbitrary molecule  $S$  with smooth boundary  $\partial S$ . The boundary separates the inner (blue) and outer (white) dielectric regions, with constants  $\epsilon_{in}$  and  $\epsilon_{out}$ , respectively. For a non-spherical DB, the distance  $r$  from the spherical center, Figure 2a, to the sampling point  $\mathbf{r}$  is replaced with the generalized expression  $r = A + p$ . Here,  $A$  is the so-called electrostatic size of  $S$ , which characterizes the dimensions of the molecule,<sup>91</sup> and  $p$  denotes the projection distance from  $\mathbf{r}$  to  $\partial S$  along the surface normal,  $\vec{n}$ .  $q_i$  is the source charge under consideration, with  $d_i$  denoting the distance between  $q_i$  and  $\mathbf{r}$ . When  $p > 0$ , the electric field normal  $\mathbf{E}_\perp(\mathbf{r})$  of  $S$  at  $\mathbf{r}$  is computed via eq 9. When  $p = 0$ , the ASC,  $\sigma(\mathbf{r})$ , is computed via eq 10.

$$\mathbf{E}_\perp(\mathbf{r}) = - \left( \frac{1}{\epsilon_{out} \left( 1 + \alpha \left( \frac{\epsilon_{in}}{\epsilon_{out}} \right) \right)} \right) \sum_i q_i \left[ \frac{\alpha \left( 1 - \frac{\epsilon_{in}}{\epsilon_{out}} \right)}{(A + p)^2} - \cos(\varphi_i) \left( \frac{(1 + \alpha)}{d_i^2} \right) \right] \quad (9)$$

where we have made the substitution  $r = A + p$  described in Figure 4. At the dielectric boundary,  $p = 0$ , applying eq 5 to eq 9 gives the following closed-form, analytical approximation for the apparent surface charge:

$$\sigma = - \left( \frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}} \right) \left( \frac{1}{4\pi \left( 1 + \alpha \left( \frac{\epsilon_{in}}{\epsilon_{out}} \right) \right)} \right) \sum_i q_i \left[ \frac{\alpha \left( 1 - \frac{\epsilon_{in}}{\epsilon_{out}} \right)}{A^2} - \cos(\varphi_i) \left( \frac{(1 + \alpha)}{d_i^2} \right) \right] \quad (10)$$

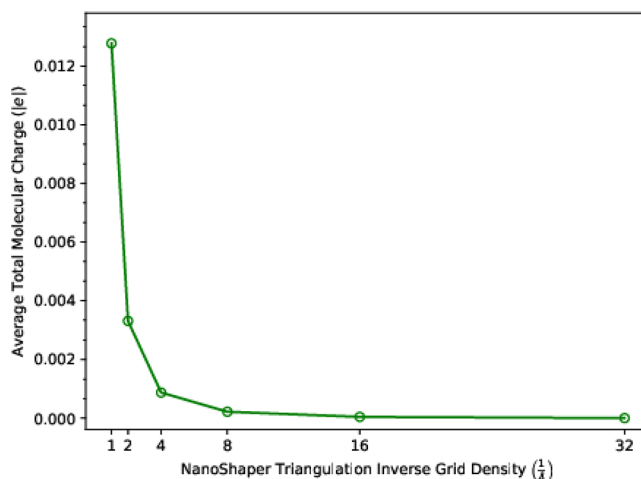
These two equations are the main analytical result of this work. Equations 7, 9, and 10 rely on approximating the  $\frac{l}{l+1}$  term of eq 4 as  $\frac{l}{l+1} = \text{const} = \alpha$  for  $l > 0$ , which allows the infinite Kirkwood series to be summed without truncation. That approach proves critical to both accuracy and computational efficiency<sup>90</sup> of the closed-form approximations, based on the Kirkwood solution. The value of  $\alpha$  in eq 7 was rigorously derived previously<sup>92</sup> to minimize RMSD to the exact Kirkwood solution for electrostatic potential, eq 4, assuming a random, discrete charge distribution inside a perfect sphere. Here, we have attempted to re-optimize the value of  $\alpha$  in the context of eq 10, aiming at best agreement with the reference for hydration free energies of our small molecule set. The effort led to only a very minor improvement in accuracy (not

shown), so we have decided to retain the original<sup>92</sup>  $\alpha = 0.580127$  for use in our ASC approximation.

**3.2.1. A Self-Consistency Check.** Arguably the simplest self-consistency check of analytical ASC is that the total surface charge produced by eq 10 should be zero for any of the neutral small molecules making up our main test set:

$$\oint_{\partial S} \sigma \, d^2s = 0 \quad (11)$$

In the discrete DB case, as the triangulation density is increased, we expect the numerical approximation of the total charge integral, eq 11, to approach zero. Thus, the same simple check automatically tests both the analytical ASC and the DB discretization (triangulation) used. As seen from Figure 5, our



**Figure 5.** Self-consistency and surface triangulation convergence check of our ASC method. The exact result for the total surface charge is zero; as NanoShaper grid density is increased, the molecular charge estimated via our implementation also tends to zero. An average over the entire set of small neutral molecules is shown. NanoShaper inverse grid spacings are given in  $\text{\AA}^{-1}$ , while average total molecular charges are given in  $|e|$ .

ASC implementation follows the expected trend. We stress that the only purpose of this simple test is a “sanity check” of our code implementation; the accuracy of the derived approximation was tested thoroughly in the sequel.

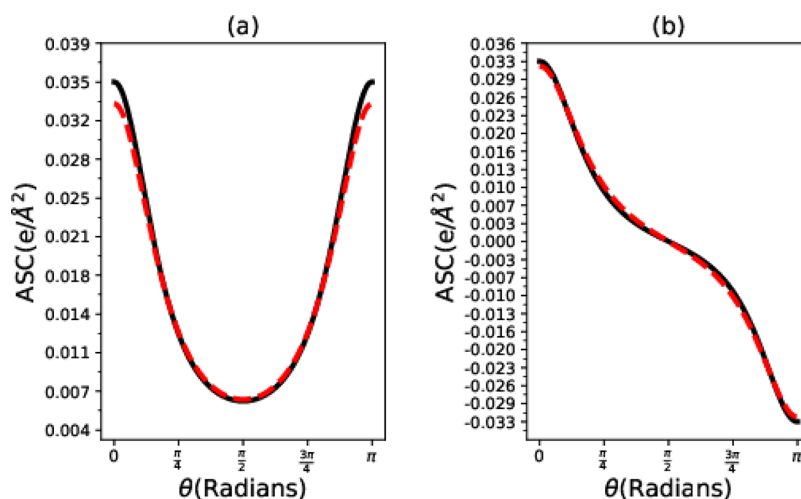
### 3.3. Accuracy against the Analytical PB Reference.

From Figure 6, our approximation matches the essentially exact Kirkwood solution (EPB), the truncation of eq 6, quite well on the two-charge test distribution, with only a slight, and expected, drop in accuracy closest to the point charges. For  $\theta \in [0, \pi]$ , an RMSD of 0.00124 and 0.00117  $e/\text{\AA}^2$  was achieved between our ASC approximation and the EPB reference, on the geometries described in panels (a) and (b) of Figure 6, respectively.

### 3.4. Electrostatic Solvation Free Energy with ASC.

The apparent surface charge formulation allows us to gain insights into a variety of solvation effects,<sup>28,93</sup> including the estimate of hydration free energy, which we used extensively here to evaluate the accuracy of our new approach against the accepted reference. Within the implicit solvation framework,<sup>8</sup> the hydration free energy of a molecule is often approximated<sup>94</sup> as the sum of polar (electrostatic) and nonpolar components:

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{np} \quad (12)$$



**Figure 6.** Apparent surface charge of our ASC approximation (dashed red) and the EPB reference (solid black) on the two-charge test distribution, shown in Figure 2b. (a) Surface charge due to two point charges, 1.5 Å away from the boundary of a dielectric sphere of radius 3 Å, with  $q_1 = q_2 = -0.65$ . These specific parameters are intended to mimic two oxygen atoms in a small molecule, with respect to the distance between a highly charged “surface” atom and the DB. (b) The same charge distribution as in panel (a) but with  $q_1 = -0.65$ ,  $q_2 = +0.65$ . Points were sampled from 0 to  $\pi$  in 0.0001 radian steps.

Of the two components in eq 12, the electrostatic solvation free energy,  $\Delta G_{el}$ , often contributes the most to the total in polar solvents such as water, especially for macromolecules. Highly approximate yet computationally efficient ways to estimate the nonpolar component,  $\Delta G_{np}$ , are widely used; by comparison,  $\Delta G_{el}$  is relatively expensive to estimate computationally with commonly used numerical methods such as NPB.<sup>7</sup> Here, our focus is  $\Delta G_{el}$ . For a discrete charge density indexed by  $i$ , we can compute  $\Delta G_{el}$  as<sup>7</sup>

$$\Delta G_{el} = \frac{1}{2} \sum_i q_i [V(\mathbf{r}_i) - V(\mathbf{r}_i)_{vac}] \quad (13)$$

where  $V(\mathbf{r}_i)$  and  $V(\mathbf{r}_i)_{vac}$  are the electrostatic potentials due to the given charge distribution in the solvent and in vacuum, respectively, sampled at each point charge  $q_i$  located at  $\mathbf{r}_i$ . In the special case when the inner and outer dielectric constants,  $\epsilon_{in}$  and  $\epsilon_{out}$  are equal to 1 and 80, respectively, we call  $\Delta G_{el}$  the electrostatic hydration free energy. We use eqs 2 and 13 to write

$$\Delta G_{el} = \frac{1}{2} \sum_i q_i \left( \oint_{\partial S} \frac{\sigma(\mathbf{s})}{|\mathbf{r}_i - \mathbf{s}|} d^2s \right) \quad (14)$$

Though eq 14 is valid for any choice of DB, the surface integral is nontrivial to compute. We approximate the surface integral using a specific triangulation of the DB, see Section 2.2. This discrete representation approximates eq 14 as

$$\Delta G_{el} \approx \left[ \sum_i \sum_T \frac{q_i \sigma_T A_T}{|\mathbf{r}_i - \mathbf{r}_T|} \right] \quad (15)$$

where  $\sigma_T$ ,  $A_T$ , and  $\mathbf{r}_T$  are the apparent surface charge, area, and center of the triangle  $T$  (to express  $\Delta G_{el}$  in kcal/mol, which is often convenient, eq 15 is multiplied by 166, while using atomic units of length Å and charge  $le$ ). The ASC on  $T$  is found by averaging the ASC at its comprising vertices. The triangular center is simply the centroid of  $T$ , with its area calculated using Heron’s formula:

$$A = \sqrt{d \cdot (d - a) \cdot (d - b) \cdot (d - c)}; \quad d = \frac{a \cdot b \cdot c}{2} \quad (16)$$

where  $a$ ,  $b$ , and  $c$  are the side lengths of  $T$ .

## 4. NUMERICAL APPLICATIONS AND RESULTS

**4.1. Analytical ASC Computational Speed.** Here, we present general running time descriptions for each tested method, rather than exact time values. In this way, we can differentiate between each method, without worrying about particular optimizations and expert parameter setups that can be found across a variety of implementations.<sup>95</sup>

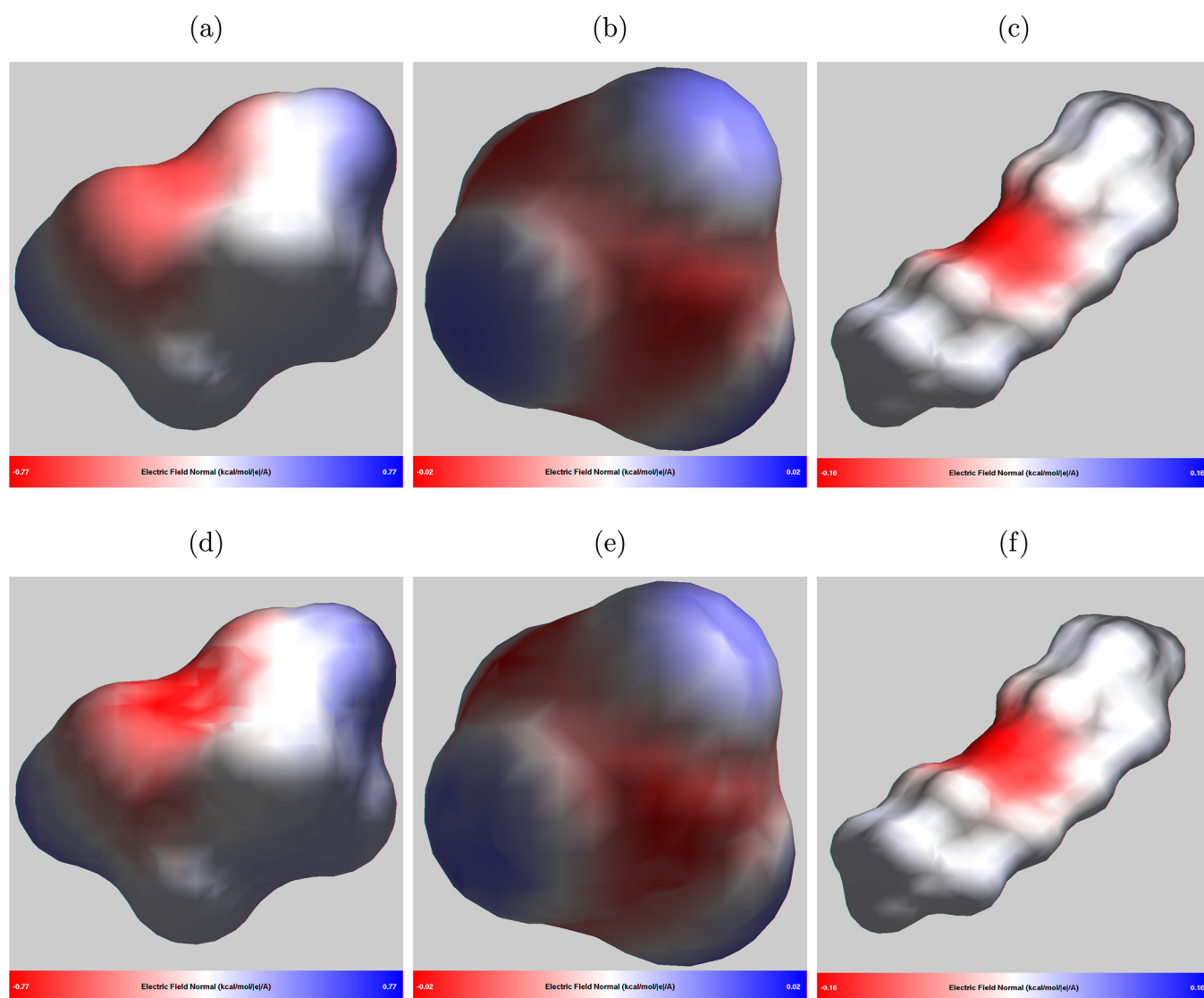
In algorithmic time complexity, the three methods we compared in Table 1 are very different. GB methods, such as

**Table 1. Running Time Expectations for Computed Electrostatic Solvation Free Energies<sup>a</sup>**

method	small molecules (average of 16 atoms)	2LZT (1958 atoms)	DNA (1598 atoms)
IGB5(AMBER)	milliseconds	~ a second	~ half a second
analytical ASC approximation	~100 ms	tens of seconds	~ half a minute
numerical PB	tens of seconds	minutes	tens of minutes

<sup>a</sup>Times are given per-molecule (averaged over the entire set in the small molecule case).

the IGB5 reference, scaled quadratically in the number of atoms  $O(K^2)$ , while our method grew linearly  $O(KN)$  in the number of atoms  $K$  and surface elements  $N$ . Volumetric methods, similar to the NPB reference, scale cubically in the number of grid points per side of a corresponding bounding box, itself a function of grid density and the maximum intramolecular distance. The impact of these asymptotic time complexities can be clearly seen when we focus on hen-egg lysozyme (2LZT) and double-stranded DNA wall running times. Though the 2LZT structure has about 400 more atoms than the DNA structure, the intramolecular width of the DNA structure is almost double that of 2LZT. This means that the



**Figure 7.** Electric field normals computed on a selection of small molecules by our ASC approximation (top row) and the NPB reference (bottom row), with visualization by GEM.<sup>36</sup> From left to right, the three molecules shown (chosen to represent various shapes) are 1,2-ethanediol (a,d), methane (b,e), and 1-butoxybutane (c,f). All calculations are made 0.7 Å from the DB, with a water probe radius of 1.4 Å. Our ASC approximation and the NPB reference use a 0.1 Å triangulation density/grid spacing. The color range for the analytical ASC and the NPB reference is the same for each vertical pair of panels.

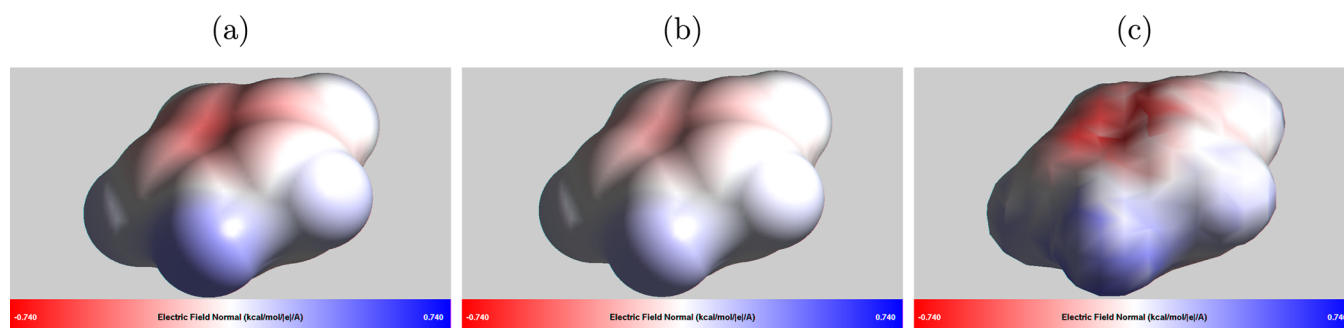
DNA structure has both a larger total surface area and requires a bigger volumetric bounding box; as seen in Table 1, we found longer running times for our ASC approximation and the NPB reference for 2LZT as compared to DNA, but not for the IGB5 reference. Hence, this contrasting algorithmic complexity affects computational timings between each model for structures of different sizes and characteristics. A principal consequence is the following: the best case scenario for the efficiency of our ASC approximation is for structures having many atoms but a comparatively low surface area; in terms of the derivation of our model, it is coincidental that in three dimensions, the shape maximizing total inner “volume” (number of atoms) and minimizing outer surface area is that of a sphere.

Though the efficiency of our analytical ASC implementation is not at the level of the IGB5 reference, it occupies a different niche: its main purpose is the estimation of the ASC and the electric field. It is worth noting that the surface integration required for eq 15 is trivially parallelizable since eqs 9 and 10

can be computed independent of adjacent surface elements on the DB. This is in addition to the more fundamental parallelism present in eqs 9 and 10, due to the independence of per-charge contributions to the electric field and ASC, respectively. Improvements in the efficiency of our implementation would greatly improve performance as surface resolution is increased, or in “worst-case” scenarios such as those seen in the DNA fragment, Table 1.

#### 4.2. Analytical ASC Accuracy with Respect to the NPB Reference.

**4.2.1. Small Molecules.** We first examine how our ASC approximation compares to the NPB reference qualitatively followed by multiple quantitative tests. For qualitative, visual comparisons, our motivation is the well-recognized utility<sup>6,9,77</sup> of visualizing electrostatic characteristics of molecules, including mapping them onto a molecular surface. Our main metric, in this section, is qualitative similarity, or the lack thereof, between visualizations produced by our analytical ASC method and the NPB reference.



**Figure 8.** Electric field normals computed via our analytical ASC method (a), the Coulomb field approximation (CFA) (b), and the NPB reference (c) on a member of our small molecule dataset, pyrrole. While both the analytical ASC and the CFA identify the positive and negative surface charge patches, the CFA underestimates their intensity significantly. Between the analytical ASC method and the CFA, minimum, maximum, and average electric field normals on the DB are  $-0.441, 0.314$ , and  $-0.030$  kcal/(mol  $\cdot$  e  $\cdot$  Å) and  $-0.281, 0.200$ , and  $-0.019$  kcal/(mol  $\cdot$  e  $\cdot$  Å), respectively. The color range used to visualize the field is the same in all the panels.

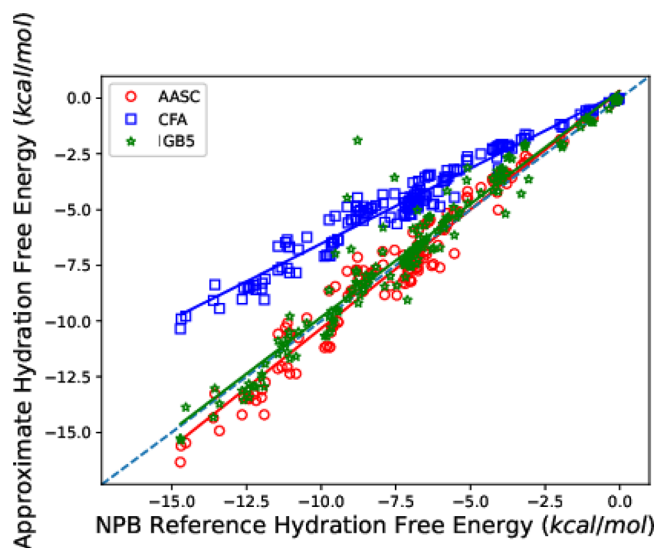
Figure 7 exemplifies the qualitative match between our method and the NPB reference on a subset of the small molecule dataset. Apart from very small irregularities attributable to the discrete sampling of the NPB potential map, the electric field normals computed with our ASC approximation are, visually, almost indistinguishable from those computed with the NPB reference.

The near quantitative agreement between the analytical ASC and the NPB reference, Figure 7 is nontrivial, and goes beyond the visual match of the “reds” and the “blues” with the NPB reference. Notice that, when  $\alpha$  is set to 0 in eqs 7 and 9, the analytical ASC reduces to the so-called Coulomb field approximation (CFA), which is often used, and can be considered a null model here.

Figure 8 clearly indicates the decreased magnitude of CFA-generated electric field normals, when compared to our ASC approximation; on average, the CFA produced an electric field that was approximately 36% weaker than our analytical ASC method, which reproduces the NPB reference closely. The significant deviation of the CFA surface charge from the reference translates into its poor accuracy in the estimation of the hydration free energy, see Figure 9.

**4.2.1.1. Quantitative Assessment of ASC Accuracy.** Next, we examine how our ASC approximation compares quantitatively to the NPB reference. Though electric field normals (or linearly related surface charges) are not the most intuitive accuracy metrics, these quantities are those that our method directly computes, and so a direct comparison with the NPB is in order. In Section 4.2.3, we discuss a physical interpretation of the deviation of these quantities from the reference. Relative to the NPB reference, our ASC approximation achieved an average RMSD and absolute difference of 0.14 and 0.11 kcal/(mol  $\cdot$  e  $\cdot$  Å) on calculated electric field normal values, respectively.

Relative to direct comparisons of molecular ASC or its proxy, the normal electric field, tests featuring the calculation of electrostatic solvation free energies are valuable in the sense that they provide an intuitive accuracy metric, directly relevant to experiments. Here, IGB5<sup>51</sup> is an example of what can be expected from a very fast GB model on small molecule datasets,<sup>96,97</sup> in terms of accuracy and running time. We tested how our ASC approximation and the IGB5 method compare to calculation of electrostatic solvation free energies by the NPB reference.



**Figure 9.** Analytical ASC (red circles), the GB (green stars), and CFA (blue squares) hydration free energies of rigid small molecules relative to the NPB reference values. Linear regression lines are plotted using colors corresponding to the representative data points. The dotted blue line represents a perfect match between an approximation and the NPB reference.  $R^2$  values for IGB5, our analytical approximation, and the CFA are 0.932, 0.961, and 0.777, respectively. Hydration free energies are shown in kcal/mol.

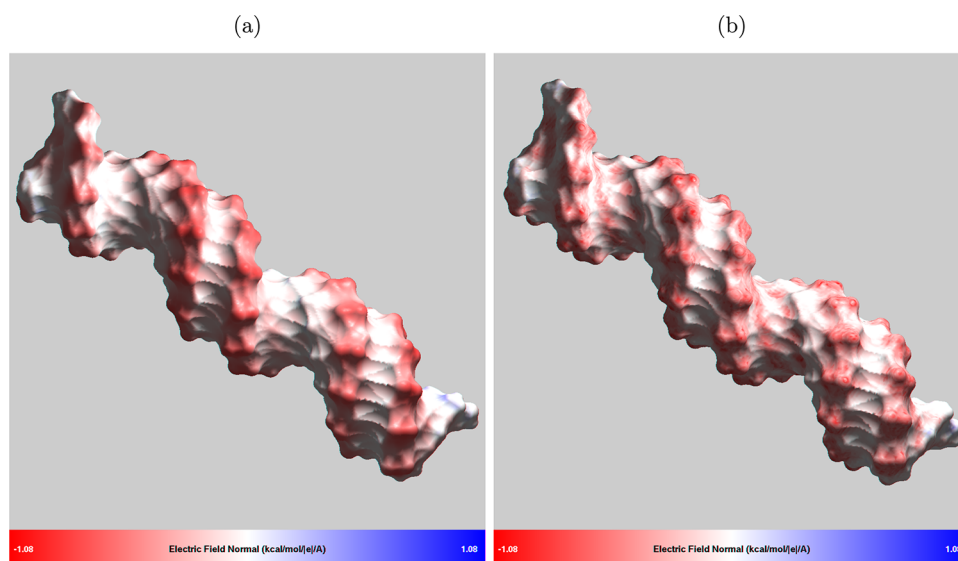
From our findings (see Table 2), our ASC approximation had an approximately 20% reduced RMSD to NPB reference hydration free energies than the IGB5 reference. This accuracy gain is encouraging, particularly from the perspective of design: GB models are designed to analytically approximate the

**Table 2. Accuracy of Electrostatic Hydration Free Energies against the NPB Reference, Estimated by our Analytical ASC Approximation, IGB5, and the CFA<sup>a</sup>**

method	RMSD to NPB reference (kcal/mol)
analytical ASC approximation	0.77
GB (IGBS,AMBER)	0.98
CFA	2.73

<sup>a</sup>Analytical  $\Delta G_{el}$  is computed via eq 14, using the analytical ASC approximation given in eq 10. Small-molecule electrostatic hydration free energies range from  $-0.01$  to  $-14.71$  kcal/mol.





**Figure 10.** Electric field normals computed on the double-stranded DNA snapshot by our ASC approximation (a) and the NPB reference (b) with visualization by GEM.<sup>36</sup> The field is estimated to be 1.5 Å from the DB, obtained with the water probe of radius of 2 Å. The larger probe radius is used here to achieve a better visualization of negative curvature regions. Our ASC approximation and the NPB reference use a 0.5 Å triangulation density/grid spacing.

electrostatic solvation free energy obtained from solving the Poisson equation. That our model can estimate solvation energies more accurately than a widely used GB model suggests that the approximation of the ASC and electric field normals by the model is reasonably accurate to be considered for practical applications. Additionally, the results in Table 2 give another encouraging conclusion, with respect to running times efficiencies. To achieve a deviation from the NPB reference of just slightly above  $kT$ , we do not require an overly fine triangulation density for the analytical ASC. When the grid resolution is set to what was used in the timing section (see Table 1), RMSD against the NPB reference differs only in nonsignificant digits. Thus, our analytical ASC can achieve a very similar accuracy, without incurring a heavy 1–2 order of magnitude time penalty, as seen with the NPB reference at this fine grid resolution.

**4.2.2. Proteins and DNA.** In analyzing the performance of our model on structures of increased size and complexity, we examine a fragment of double-stranded DNA and the hen-egg lysozyme. Structures of this type—with regions of the DB having deep, negative curvature “pockets”—present some of the toughest tests for our model due to certain theoretical considerations we touch on in section 4.2.3.2.

**4.2.2.1. Double-Stranded DNA.** First, we examined our analytical approximation on a double-stranded DNA fragment.

Qualitatively, Figure 10 shows that our analytical approximation reproduced the NPB reference quite well, although some discrepancies are clearly seen in the grooves, that is, in regions of high negative curvature.

**4.2.2.2. Triclinic Hen Egg White Lysozyme.** Next, we compared our analytical ASC to the NPB reference on the triclinic hen egg white lysozyme.

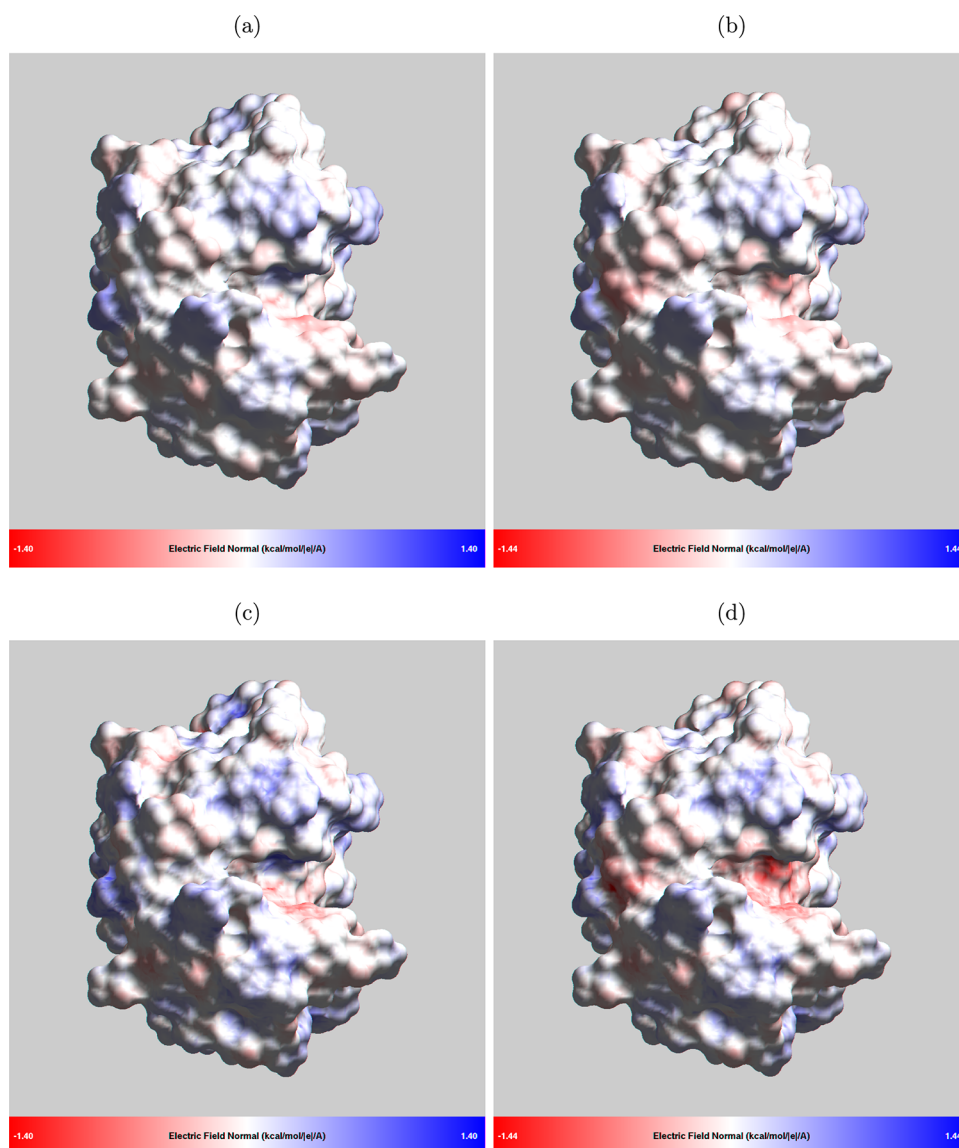
We see in Figure 11 that our analytical approximation accurately reproduced the NPB reference, outside of the hen-egg lysozyme’s binding cleft. Within the binding cleft, quantitative deviations in electric field normal magnitudes from the NPB reference become apparent, though our approximation still produces a qualitatively reasonable picture.

Our approximation (Figure 11a,b) qualitatively reproduced the reference (Figure 11c,d) in visualizing the substantial electrostatic effect of Asp 52 and Glu 35 in the enzymatic pocket as well as the corresponding changes due to the change in the charge states of these two residues under mildly acidic conditions (pH 4.5). Hence, both the NPB reference and our model are able to visualize the behavior of Asp 52 and Glu 35, under a pH change.<sup>98–100</sup>

**4.2.2.3. Quantitative Assessment of ASC Accuracy.** With qualitative tests complete, we finish the analysis with a quantitative comparison between our analytical approximation and the NPB reference.

As expected, quantitative performance deficiencies exist for larger molecules with prominent regions of negative curvature. Although the double-stranded DNA and hen-egg lysozyme have similar numbers of atoms, average RMSD values in Table 3 are, relatively, inconsistent. On the double-stranded DNA snapshot ( $\sim 1600$  atoms), the average RMSD against the NPB reference is about 2.6 times larger than on small molecules (Section 4.2.1, Table 3). Comparatively, the average RMSD of the hen-egg lysozyme ( $\sim 2000$  atoms) is only about 1.4 times larger than on small molecules. On hydration free energies, relative errors between our analytical approximation and the NPB reference are quite small on the DNA snapshot,  $\sim 4\%$ , but more than double,  $\sim 12\%$ , on the hen-egg lysozyme. These findings might have to do with the DNA’s proportion of negative curvature regions with respect to the whole.

**4.2.3. Discussion.** **4.2.3.1. Surface Charge Distribution and Electrostatic Solvation Free Energy.** To put the results of Section 4.2.1 and Table 3 in a better context, it may be helpful to consider a hypothetical situation. Suppose a biomolecule of interest has a constant electric field strength near its dielectric boundary. When compared to the reference, there is, on average, a  $\sim 0.4$  kcal/(mol  $\cdot e \cdot \text{Å}$ ) RMSD error in the electric field normal values. If a unit electric charge is moved 1 Å away from the biomolecular boundary, along the surface normal, the error in the total work done by the electric field would be less than  $\sim 0.4$  kcal/mol—small when compared to the “gold-



**Figure 11.** Electric field normals computed on the hen-egg lysozyme by our ASC approximation (top row) and the NPB reference (bottom row), with visualization by GEM.<sup>36</sup> (a,c) Structure at pH 4.5. (b,d) Structure at pH 6.5. All calculations are made 1.5 Å from the DB, with a water probe radius of 2 Å. The larger probe radius is used to achieve a better visualization of negative curvature regions. Our ASC approximation and the NPB reference use a 0.5 Å triangulation density/grid spacing.

**Table 3. Electric Field Normal Comparisons between Our Analytical Approximation and the NPB Reference, on Double-Stranded DNA and the Protonated/Unprotonated Hen-Egg Lysozyme<sup>a</sup>**

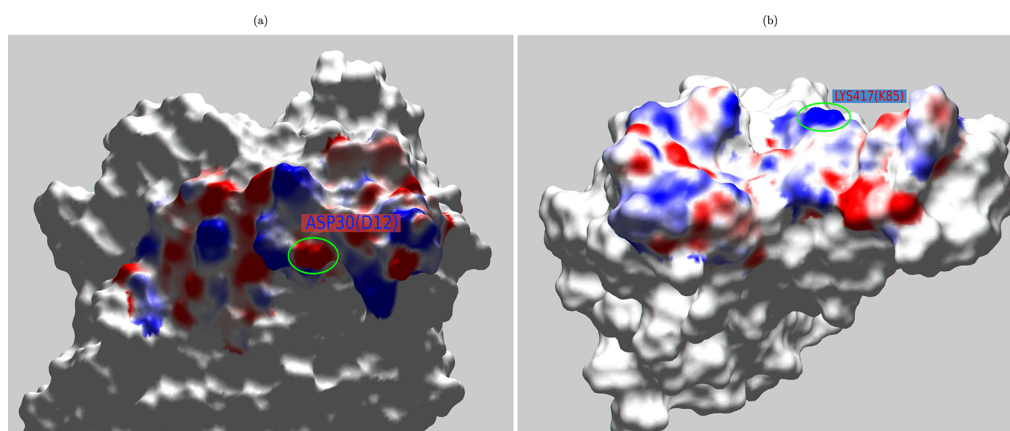
	double-stranded DNA	2LZT pH 4.5	2LZT pH 6.5
absolute difference	0.27	0.15	0.15
average RMSD	0.37	0.19	0.20

<sup>a</sup>All values are in kcal/(mol · e · Å).

standard” 1 kcal/mol difference against reference. Several caveats are due here. First, the estimate is based on the RMS error; it is possible that the errors are larger in some spots near the DB. That concern is mitigated to an extent by the fact that the electric field strength is inversely related to the square of distance from the surface and would, in reality, decrease away from the surface, making any discrepancy with the reference smaller. Second, this test is free from the danger of a specific

error cancellation that may affect the commonly used error assessment based on solvation free energies. As the history of the GB model development demonstrates,<sup>41</sup> very large errors in individual contributions may cancel out to produce seemingly accurate total solvation energies.

**4.2.3.2. Structural Considerations.** A notable feature, seen prominently in both the double-stranded DNA snapshot and the hen-egg lysozyme, but not generally in small molecules, is the presence of distinct, and fairly deep, negative curvature pockets on the DB. Our analytical ASC model is derived from an exact solution of the Poisson problem on a spherical DB (Figure 2a), having positive curvature throughout. Negative curvature regions, such as the main groove in Figure 10 and the binding cleft in Figure 11, do not occur on a sphere; this is where our model is not expected to perform well. A resulting loss in performance had been noted previously<sup>36</sup> for the approximate electrostatic potential (eq 7). It is therefore surprising that a qualitative agreement with the NPB reference



**Figure 12.** Apparent surface charge computed on the receptor binding domain (RBD) of ACE2 receptor/SARS-CoV-2 spike glycoprotein complex, with visualization by GEM.<sup>36</sup> The range of the ASC values is  $\pm 0.008 e/\text{\AA}^2$ , corresponding to the red-blue color map. The white area—a majority of the molecular surface outside of the RBD—is excluded from the calculation, which reduces the computational time significantly. To mimic the color convention used in Figure 1B of Wang et al.<sup>77</sup> the sign of eq 10 is reversed for this calculation; that is, the negative of the  $\sigma$  is shown. (a,b) ACE2 receptor and SARS-CoV-2 spike glycoprotein RBDs, respectively. All calculations are made 0.7 Å from the DB with a water probe radius of 2 Å. A NanoShaper triangulation density of 0.5 Å is used.

is still seen, even in these regions. More shallow negative curvature regions are also present in the small molecules, Figure 7, but apparently these do not present a significant challenge to the new model.

Confounding this effect, deep negative curvature regions on the DB can restrict water molecule conformational freedom, making nearby solvent behave less similarly to that of the bulk.<sup>101,102</sup> It has been shown that regions of this type can significantly modify interactions between small molecule inhibitors and their target proteins,<sup>103</sup> prompting investigations related to their identification.<sup>104,105</sup> In our context, this change in the behavior of water bulk has negative implications on the performance of our model, but the same is true for the NPB reference, which is also based on the continuum solvent. Because both of these models are expected to deviate from the correct physical behavior within these regions of negative curvature, we argue that a qualitative agreement with the NPB reference may be acceptable here, in place of a strong quantitative agreement.

**4.2.3.3. Computational Considerations.** Though not implemented in this work, one very important consequence of our analytical approach to the ASC confers a key additional theoretical benefit—the trivially parallel nature of our solution. Equation 10 is computed over our discrete DB representation and is totally independent from surrounding surface elements. Coupled with a similarly parallelizable expression for the computation of electrostatic solvation free energy, eq 15, our method holds significant potential for the computationally efficient ASC treatment of large biomolecules.

**4.3. Testing on a Large Biomolecule.** For a real-world application, we examine our approximation on a much larger (~6500 atom) complex, with important relevance today—the ACE2/SARS-CoV-2 complex (PDB ID: 6M0J).

Recently, a comparison has been made between SARS-CoV and SARS-CoV-2, examining various mutations and their effects on respective binding strengths with the ACE2 agonist.<sup>77</sup> The focus of Figure 12 is on, what Wang et al.<sup>77</sup> termed, the “CR2” receptor binding domain; the visualized ASC shows how our approximation reproduced the electrostatic complementary of surfaces charges between two residues, ASP30 (D12), Figure 12a, and LYS 417 (K85),

Figure 12b, thought to contribute to the formation of a salt bridge. This salt bridge improves both stability and binding strength between the ACE2 receptor and SARS-CoV-2 spike protein, when compared to the SARS-CoV spike protein.

These large-scale visualizations of the ASC (or of the normal component of the electric field) has already been shown useful;<sup>6</sup> our analytical approximation might be a useful tool in understanding complex protein–protein interactions at an atomistic scale, including SARS-CoV-2 mutants of concern,<sup>106</sup> especially in high throughput studies. Our method can potentially be useful in this area due to its targeted, source-based approach to computation of ASC, where only a small portion the entire DB, and hence a small subset of the surface elements, is included in the computation, as demonstrated in Figure 12. As a result, the computational time is reduced dramatically—see the discussion on time complexity in section 4.2.3.3.

## 5. CONCLUSIONS

In this work, we have derived a closed-form, analytical approximation for biomolecular apparent surface charge (ASC) and the normal component of the electric field outside the molecule. To the best of our knowledge, this is the first such fully analytical approximation. The approximation was constructed to closely reproduce the exact infinite series solution for a perfect spherical boundary; more importantly, the approximation yields a reasonably close agreement with the standard numerical PB for realistic molecular structures. Specifically, quantitative reproduction of results from the standard numerical PB reference was achieved on most of the tested molecules, except within prominent regions of negative curvature, where the new approximation is still qualitatively correct. Comparisons with a popular fast GB model in AMBER (IGBS) shows that our method is more accurate in reproducing the hydration free energy, albeit at higher computational expense, which may be expected of a proof-of-concept code package that is not highly optimized. At the same time, standard numerical PB is still 1–2 orders of magnitude slower than the proposed approximation, which puts it “in-between” fast analytical GB and numerical PB. We stress that solvation free energy estimates are used here as a

common and convenient accuracy metric and is not where we believe the potential benefits of the proposed analytical ASC may be. These potential benefits stem from the unique features of the method. In our view, the main advantage of ASC compared to the GB is its versatility: in contrast to the GB, ASC allows one to estimate pretty much any electrostatic quantity.

There are at least two features of the new approximation absent from the GB: the ability to estimate the apparent surface charge (and, hence, the potential everywhere) and the ability to estimate the normal component of the electric field. Another noteworthy feature of the approach sets it apart from other existing approximations that can estimate ASC, including those aimed at computing ASC directly—the fact that the new approximation is “source-based”. This means that the normal electric field and the ASC can be estimated at any individual point or surface patch, without the need for self-consistent computation over the entire surface or volume. This feature is in contrast to “field-based” methods such as numerical solutions of the Poisson equation or DPCM. As an illustration, we showed that the “source-based” feature of our ASC approximation allows a rapid examination of the ACE2/SARS-CoV-2 RBD electrostatics, reproducing conditions posited to contribute to the spike protein’s high binding strength.

An area which, in our view, can benefit the most from the proposed analytical ASC is the development of new implicit solvation methods that require fast estimates of local polarization charges and/or fields. We also believe that the new approach may have the potential to compete with existing ASC-based approaches in QM applications, especially where computational efficiency is key; further extensive testing and analysis beyond the scope of the proof-of-concept work will be necessary to explore the potential of the method in this area.

As it stands, the proposed method has several limitations. First, it does not yet include salt effects explicitly. However, in the future, it should be relatively easy to add into the model salt dependence at the Debye–Hückel level, following an approach outlined in ref 36. Another limitation of the model is its qualitative nature in the regions of high negative curvature, at least relative to the standard NPB reference. Overcoming this specific limitation will require a significant extension of the underlying theory and extensive testing on biomolecular structures.

A careful and detailed comparison of our proof-of-concept approximation within the broader category of existing, optimized implementations ASC methods has not been performed and is warranted in the future; nonetheless, promising results so far have pointed to the potential of our approach in forming the basis of novel implicit models of solvation.

## AUTHOR INFORMATION

### Corresponding Author

Alexey V. Onufriev – Department of Computer Science, Department of Physics, and Center for Soft Matter and Biological Physics, Virginia Tech, Blacksburg, Virginia 24061, United States; [orcid.org/0000-0002-4930-6612](https://orcid.org/0000-0002-4930-6612); Email: [alexey@cs.vt.edu](mailto:alexey@cs.vt.edu)

### Author

Dan E. Folescu – Department of Computer Science, Virginia Tech, Blacksburg, Virginia 24061, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.2c01484>

## Notes

The authors declare no competing financial interest.

A.V.O. reports support from the NIH, grants R21 GM131228 and R01 GM144596.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ACKNOWLEDGMENTS

We would like to thank Nitin Passa for his contributions to the numerical ideas behind this application of ASC. In addition, we appreciate the help of Dr. Igor Tolokh for supplying the double-stranded DNA snapshot and Dr. Andrew Fenley for his useful comments and suggestions.

## REFERENCES

- (1) Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* **2018**, *118*, 1691–1741.
- (2) Adcock, S. A.; McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev.* **2006**, *106*, 1589–1615.
- (3) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
- (4) Raman, E. P.; MacKerell, A. D. Spatial Analysis and Quantification of the Thermodynamic Driving Forces in Protein-Ligand Binding: Binding Site Variability. *J. Am. Chem. Soc.* **2015**, *137*, 2608–2621.
- (5) Amaro, R. E.; Mulholland, A. J. A Community Letter Regarding Sharing Biomolecular Simulation Data for COVID-19. *J. Chem. Inf. Model.* **2020**, *60*, 2653–2656.
- (6) Tolokh, I. S.; Pabit, S. A.; Katz, A. M.; Chen, Y.; Drozdetski, A.; Baker, N.; Pollack, L.; Onufriev, A. V. Why Double-Stranded RNA Resists Condensation. *Nucleic Acids Res.* **2014**, *42*, 10823–10831.
- (7) Onufriev, A. V.; Izadi, S. Water Models for Biomolecular Simulations. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, *8*, No. e1347.
- (8) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.
- (9) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268*, 1144–1149.
- (10) Beroza, P.; Case, D. A. Calculations of Proton-Binding Thermodynamics in Proteins. *Methods Enzymol.* **1998**, *295*, 170–189.
- (11) Madura, J. D.; Davist, M. E.; Gilson, M. K.; Wades, R. C.; Luty, B. A.; McCammon, J. A. *Reviews in Computational Chemistry*; John Wiley & Sons, Ltd, 1994, pp. 229–267.
- (12) Gilson, M. K. Theory of Electrostatic Interactions in Macromolecules. *Curr. Opin. Struct. Biol.* **1995**, *5*, 216–223.
- (13) Scarsi, M.; Apostolakis, J.; Cafilisch, A. Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions. *J. Phys. Chem. A* **1997**, *101*, 8098–8106.
- (14) Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson–Boltzmann Calculations for Static and Dynamic Systems. *J. Comput. Chem.* **2002**, *23*, 1244–1253.
- (15) Simonson, T. Electrostatics and Dynamics of Proteins. *Rep. Prog. Phys.* **2003**, *66*, 737–787.
- (16) Baker, N. A.; Bashford, D.; Case, D. A. In *New Algorithms for Macromolecular Simulation*; Leimkuhler, B., Chipot, C., Elber, R., Laaksonen, A., Mark, A., Schlick, T., Schütte, C., Skeel, R., Eds.; Lecture Notes in Computational Science and Engineering; Springer: Berlin, Heidelberg, 2006, pp. 263–295.
- (17) Bardhan, J. P. Biomolecular Electrostatics—I Want Your Solvation (Model)\*. *Comput. Sci. Discov.* **2012**, *5*, No. 013001.
- (18) Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric “Constant” of Proteins: Smooth Dielectric Function for Macro-

molecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.

(19) Rocchia, W.; Alexov, E.; Honig, B. Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J. Phys. Chem. B* **2001**, *105*, 6507–6514.

(20) Nicholls, A.; Honig, B. A Rapid Finite Difference Algorithm, Utilizing Successive over-Relaxation to Solve the Poisson–Boltzmann Equation. *J. Comput. Chem.* **1991**, *12*, 435–445.

(21) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037–10041.

(22) Chen, D.; Chen, Z.; Chen, C.; Geng, W.; Wei, G.-W. MIBPB: A Software Package for Electrostatic Analysis. *J. Comput. Chem.* **2011**, *32*, 756–770.

(23) Nguyen, D. D.; Wang, B.; Wei, G.-W. Accurate, Robust, and Reliable Calculations of Poisson–Boltzmann Binding Energies. *J. Comput. Chem.* **2017**, *38*, 941–948.

(24) Jackson, J. D. *Classical Electrodynamics*, 3rd ed., Wiley: New York, 1999.

(25) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilizaion of AB Initio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* **1981**, *55*, 117–129.

(26) Miertus, S.; Tomasi, J. Approximate Evaluations of the Electrostatic Free Energy and Internal Energy Changes in Solution Processes. *Chem. Phys.* **1982**, *65*, 239–245.

(27) Onufriev, A. *Modeling Solvent Environments*; John Wiley & Sons, Ltd, 2010; Chapter 6, pp. 127–165.

(28) Onsager, L. Electric Moments of Molecules in Liquids. *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.

(29) Barker, J.; Watts, R. Monte Carlo Studies of the Dielectric Properties of Water-like Models. *Mol. Phys.* **1973**, *26*, 789–792.

(30) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224–2235.

(31) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3094.

(32) Herbert, J. M. Dielectric Continuum Methods for Quantum Chemistry. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11*, No. e1519.

(33) Truong, T. N.; Stefanovich, E. V. A New Method for Incorporating Solvent Effect into the Classical, Ab Initio Molecular Orbital and Density Functional Theory Frameworks for Arbitrary Shape Cavity. *Chem. Phys. Lett.* **1995**, *240*, 253–260.

(34) Cossi, M.; Barone, V. Analytical Second Derivatives of the Free Energy in Solution by Polarizable Continuum Models. *J. Chem. Phys.* **1998**, *109*, 6246–6254.

(35) Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, *1*, 799–805.

(36) Gordon, J. C.; Fenley, A. T.; Onufriev, A. An Analytical Approach to Computing Biomolecular Electrostatic Potential. II. Validation and Applications. *J. Chem. Phys.* **2008**, *129*, No. 075102.

(37) Hoijtink, G. J.; de Boer, E.; van der Meij, P. H.; Weijland, W. P. Reduction Potentials of Various Aromatic Hydrocarbons and Their Univalent Anions. *Recl. Trav. Chim. Pays-Bas* **1956**, *75*, 487–503.

(38) Tucker, S. C.; Truhlar, D. G. Generalized Born Fragment Charge Model for Solvation Effects as a Function of Reaction Coordinate. *Chem. Phys. Lett.* **1989**, *157*, 164–170.

(39) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.

(40) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(41) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.

(42) Dominy, B. N.; Brooks, C. L. Development of a Generalized Born Model Parametrization for Proteins and Nucleic Acids. *J. Phys. Chem. B* **1999**, *103*, 3765–3773.

(43) Bashford, D.; Case, D. A. Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.

(44) Calimet, N.; Schaefer, M.; Simonson, T. Protein Molecular Dynamics with the Generalized Born/ACE Solvent Model. *Proteins* **2001**, *45*, 144–158.

(45) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* **1995**, *246*, 122–129.

(46) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.* **1996**, *100*, 19824–19839.

(47) Schaefer, M.; Karplus, M. A Comprehensive Analytical Treatment of Continuum Electrostatics. *J. Phys. Chem.* **1996**, *100*, 1578–1599.

(48) Feig, M.; Im, W.; Brooks, C. L. Implicit Solvation Based on Generalized Born Theory in Different Dielectric Environments. *J. Chem. Phys.* **2004**, *120*, 903–911.

(49) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. Novel Generalized Born Methods. *J. Chem. Phys.* **2002**, *116*, 10606–10614.

(50) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L. New Analytic Approximation to the Standard Molecular Volume Definition and Its Application to Generalized Born Calculations. *J. Comput. Chem.* **2003**, *24*, 1348–1356.

(51) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* **2004**, *55*, 383–394.

(52) Wang, T.; Wade, R. C. Implicit Solvent Models for Flexible Protein-Protein Docking by Molecular Dynamics Simulation. *Proteins* **2003**, *50*, 158–169.

(53) Gallicchio, E.; Levy, R. M. AGBNP: An Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High-Resolution Modeling. *J. Comput. Chem.* **2004**, *25*, 479–499.

(54) Ghosh, A.; Rapp, C. S.; Friesner, R. A. Generalized Born Model Based on a Surface Integral Formulation. *J. Phys. Chem. B* **1998**, *102*, 10983–10990.

(55) Im, W.; Lee, M. S.; Brooks, C. L. Generalized Born Model with a Simple Smoothing Function. *J. Comput. Chem.* **2003**, *24*, 1691–1702.

(56) Haberthür, U.; Cafilisch, A. FACTS: Fast Analytical Continuum Treatment of Solvation. *J. Comput. Chem.* **2008**, *29*, 701–715.

(57) Grant, J. A.; Pickup, B. T.; Sykes, M. J.; Kitchen, C. A.; Nicholls, A. The Gaussian Generalized Born Model: Application to Small Molecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4913–4922.

(58) Tjong, H.; Zhou, H.-X. GBr(6): A Parameterization-Free, Accurate, Analytical Generalized Born Method. *J. Phys. Chem. B* **2007**, *111*, 3055–3061.

(59) Labute, P. The Generalized Born/Volume Integral Implicit Solvent Model: Estimation of the Free Energy of Hydration Using London Dispersion Instead of Atomic Surface Area. *J. Comput. Chem.* **2008**, *29*, 1693.

(60) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect. *J. Comput. Chem.* **2002**, *23*, 1297–1304.

(61) Onufriev, A. V.; Sigalov, G. A Strategy for Reducing Gross Errors in the Generalized Born Models of Implicit Solvation. *J. Chem. Phys.* **2011**, *134*, 164104.

(62) Lange, A. W.; Herbert, J. M. Improving Generalized Born Models by Exploiting Connections to Polarizable Continuum Models. I. An Improved Effective Coulomb Operator. *J. Chem. Theory Comput.* **2012**, *8*, 1999–2011.

- (63) Simmerling, C.; Strockbine, B.; Roitberg, A. E. All-Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Am. Chem. Soc.* **2002**, *124*, 11258–11259.
- (64) Chen, J.; Im, W.; Brooks, C. L. Balancing Solvation and Intramolecular Interactions: Toward a Consistent Generalized Born Force Field. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.
- (65) Lei, H.; Duan, Y. Ab Initio Folding of Albumin Binding Domain from All-Atom Molecular Dynamics Simulation. *J. Phys. Chem. B* **2007**, *111*, 5458–5463.
- (66) Jagielska, A.; Scheraga, H. A. Influence of Temperature, Friction, and Random Forces on Folding of the B-domain of Staphylococcal Protein A: All-atom Molecular Dynamics in Implicit Solvent. *J. Comput. Chem.* **2007**, *28*, 1068–1082.
- (67) Amaro, R. E.; Cheng, X.; Ivanov, I.; Xu, D.; McCammon, J. A. Characterizing Loop Dynamics and Ligand Recognition in Human- and Avian-Type Influenza Neuraminidases via Generalized Born Molecular Dynamics and End-Point Free Energy Calculations. *J. Am. Chem. Soc.* **2009**, *131*, 4702–4709.
- (68) Ruscio, J. Z.; Onufriev, A. A Computational Study of Nucleosomal DNA Flexibility. *Biophys. J.* **2006**, *91*, 4121–4132.
- (69) Chocholoušová, J.; Feig, M. Implicit Solvent Simulations of DNA and DNA-Protein Complexes: Agreement with Explicit Solvent vs Experiment. *J. Phys. Chem. B* **2006**, *110*, 17240–17251.
- (70) Zhu, J.; Alexov, E.; Honig, B. Comparative Study of Generalized Born Models: Born Radii and Peptide Folding. *J. Phys. Chem. B* **2005**, *109*, 3008–3022.
- (71) Onufriev, A. V.; Case, D. A. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 275–296.
- (72) Mobley, D. L.; Guthrie, J. P. FreeSolv: A Database of Experimental and Calculated Hydration Free Energies, with Input Files. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 711–720.
- (73) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Correction to Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 1347–1347.
- (74) Mukhopadhyay, A.; Aguilar, B. H.; Tolokh, I. S.; Onufriev, A. V. Introducing Charge Hydration Asymmetry into the Generalized Born Model. *J. Chem. Theory Comput.* **2014**, *10*, 1788–1794.
- (75) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (76) Ramanadham, M.; Sieker, L. C.; Jensen, L. H. Refinement of Triclinic Lysozyme: II. The Method of Stereochemically Restrained Least Squares. *Acta Crystallogr., Sect. B: Struct. Sci.* **1990**, *46*, 63–69.
- (77) Wang, X.; Lan, J.; Ge, J.; Yu, J.; Shan, S. Crystal Structure of SARS-CoV-2 Spike Receptor-Binding Domain Bound with ACE2. <https://www.rcsb.org/structure/6M0J>, 2020.
- (78) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H<sup>++</sup>: A Server for Estimating pK<sub>a</sub>s and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005**, *33*, W368–W371.
- (79) Connolly, M. L. Analytical Molecular Surface Calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.
- (80) Decherchi, S.; Rocchia, W. A General and Robust Ray-Casting-Based Algorithm for Triangulating Surfaces at the Nanoscale. *PLoS One* **2013**, *8*, No. e59744.
- (81) Grant, J. A.; Pickup, B. T.; Nicholls, A. A Smooth Permittivity Function for Poisson–Boltzmann Solvation Methods. *J. Comput. Chem.* **2001**, *22*, 608–640.
- (82) Tjong, H.; Zhou, H.-X. On the Dielectric Boundary in Poisson–Boltzmann Calculations. *J. Chem. Theory Comput.* **2008**, *4*, 507–514.
- (83) Cramer, C. J.; Truhlar, D. G. A Universal Approach to Solvation Modeling. *Acc. Chem. Res.* **2008**, *41*, 760–768.
- (84) Onufriev, A. V.; Aguilar, B. Accuracy of Continuum Electrostatic Calculations Based on Three Common Dielectric Boundary Definitions. *J. Theor. Comput. Chem.* **2014**, *13*, 1440006.
- (85) Richards, F. M. Areas, Volumes, Packing, and Protein Structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176.
- (86) Bondi, A. Van Der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441–451.
- (87) Bashford, D. An Object-Oriented Programming Suite for Electrostatic Effects in Biological Molecules An Experience Report on the MEAD Project. *Scientific Computing in Object-Oriented Parallel Environments*. Berlin, Heidelberg, 1997, pp. 233–240.
- (88) Zhou, Y. C.; Feig, M.; Wei, G. W. Highly Accurate Biomolecular Electrostatics in Continuum Dielectric Environments. *J. Comput. Chem.* **2008**, *29*, 87–97.
- (89) Kirkwood, J. G. Theory of Solutions of Molecules Containing Widely Separated Charges with Special Application to Zwitterions. *J. Chem. Phys.* **1934**, *2*, 351–361.
- (90) Fenley, A. T.; Gordon, J. C.; Onufriev, A. An Analytical Approach to Computing Biomolecular Electrostatic Potential. I. Derivation and Analysis. *J. Chem. Phys.* **2008**, *129*, No. 075101.
- (91) Sigalov, G.; Fenley, A.; Onufriev, A. Analytical Electrostatics for Biomolecules: Beyond the Generalized Born Approximation. *J. Chem. Phys.* **2006**, *124*, 124902.
- (92) Sigalov, G.; Scheffel, P.; Onufriev, A. Incorporating Variable Dielectric Environments into the Generalized Born Model. *J. Chem. Phys.* **2005**, *122*, No. 094511.
- (93) Cammi, R.; Tomasi, J. Remarks on the Use of the Apparent Surface Charges (ASC) Methods in Solvation Problems: Iterative versus Matrix-Inversion Procedures and the Renormalization of the Apparent Charges. *J. Comput. Chem.* **1995**, *16*, 1449–1458.
- (94) Honig, B.; Sharp, K.; Yang, A. S. Macroscopic Models of Aqueous Solutions: Biological and Chemical Applications. *J. Phys. Chem.* **1993**, *97*, 1101–1109.
- (95) Aguilar, B.; Onufriev, A. V. Efficient Computation of the Total Solvation Energy of Small Molecules via the R6 Generalized Born Model. *J. Chem. Theory Comput.* **2012**, *8*, 2404–2411.
- (96) Katkova, E. V.; Onufriev, A. V.; Aguilar, B.; Sulimov, V. B. Accuracy Comparison of Several Common Implicit Solvent Models and Their Implementations in the Context of Protein-Ligand Binding. *J. Mol. Graphics* **2017**, *72*, 70–80.
- (97) Izadi, S.; Harris, R. C.; Fenley, M. O.; Onufriev, A. V. Accuracy Comparison of Generalized Born Models in the Calculation of Electrostatic Binding Free Energies. *J. Chem. Theory Comput.* **2018**, *14*, 1656–1670.
- (98) Blake, C. C. F.; Johnson, L. N.; Mair, G. A.; North, A. C. T.; Phillips, D. C.; Sarma, V. R.; Perutz, M. F. Crystallographic Studies of the Activity of Hen Egg-White Lysozyme. *Proc. R. Soc. London, Ser. B* **1967**, *167*, 378–388.
- (99) Phillips, D. C. The Three-Dimensional Structure of an Enzyme Molecule. *Sci. Am.* **1966**, *215*, 78–90.
- (100) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (101) Li, Z.; Lazaridis, T. Water at Biomolecular Binding Interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573–581.
- (102) de Beer, S. B. A.; Vermeulen, N. P. E.; Oostenbrink, C. The Role of Water Molecules in Computational Drug Design. *Curr. Top. Med. Chem.* **2010**, *10*, 55–66.
- (103) Spyralis, F.; Ahmed, M. H.; Bayden, A. S.; Cozzini, P.; Mozzarelli, A.; Kellogg, G. E. The Roles of Water in the Protein Matrix: A Largely Untapped Resource for Drug Discovery. *J. Med. Chem.* **2017**, *60*, 6781–6827.
- (104) Hamelberg, D.; McCammon, J. A. Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.
- (105) Petřek, M.; Otyepka, M.; Banáš, P.; Košinová, P.; Koča, J.; Damborský, J. CAVER: A New Tool to Explore Routes from Protein Clefts, Pockets and Cavities. *BMC Bioinf.* **2006**, *7*, 316.
- (106) Zahradník, J.; Marciano, S.; Shemesh, M.; Zoler, E.; Chiaravalli, J.; Meyer, B.; Dym, O.; Elad, N.; Schreiber, G. SARS-CoV-2 RBD in Vitro Evolution Follows Contagious Mutation Spread, yet Generates an Able Infection Inhibitor. *bioRxiv* **2021**, No. 425392.